Image Categorization Based on Clustering Spatial Frequency Maps

Fuhui Long [1], Hanchuan Peng [2], and David Dagan Feng [3,4]

[1] *Duke University Medical Center, Duke University*
Office: Box 90999, LSRC Bldg., Room B243E, Research Drive
Duke University, Durham, North Carolina 27708, USA
Email: long@neuro.duke.edu.

[2] *Computer Science and Mathematics Division, Oak Ridge National Lab.*
Office: A110 Life Science Building, 120 Green St., Athens, GA, 30602.
Email: hpeng@csbl.bmb.uga.edu.

[3]*Department of Electronic and Information Engineering, The Hong Kong Polytechnic University*
Hung Hom, Kowloon, Hong Kong. Email: enfeng@polyu.edu.hk

[4]*School of Information Technologies, University of Sydney, NSW 2006, Australia*
Email: feng@it.usyd.edu.au

ABSTRACT

Image classification can facilitate semantic retrieval and browsing of large-scale image databases. Existing approaches are usually based on extracting local or global low-level features such as color, edge, and texture from images. In this paper, we propose an image categorization method that characterizes the respective scene structures in images. 2D Spatial Frequency Map of an image, as well as the respective projection vector representations and principal component representations, are used to characterize the spatial structure of the image. Based on multiple similarity scores, we use a spectral clustering method and a maximal-spanning-tree-spectral-clustering method to generate image categories.

Keywords: image categorization; spatial frequency; spectral clustering; maximal spanning tree

1. INTRODUCTION

Automatic image categorization can facilitate many visual processing tasks such as scene recognition and large-scale image database retrieval/browsing. Most existing methods tackle this problem by clustering images based on various local or global low-level visual features such as edge, color and texture [5][11][15]. Some techniques such as self-organization map, hierarchical classification, and Bayesian networks have been proposed to correlate low-level features with higher-level semantic meaning of images.

Despite their successes, most existing methods focus on two-class problems (e.g. indoor/ outdoor, or city/landscape). In addition, usually they do not consider specifically the global spatial structure information embedded in the images. For instance, images of lakes in the long distance may contain several horizontally distributed smooth areas, corresponding to the sky at the top,

Storage and Retrieval Methods and Applications for Multimedia 2004, edited by Minerva M. Yeung, Rainer W. Lienhart, Chung-Sheng Li, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 5307 © 2004 SPIE and IS&T · 0277-786X/04/$15

buildings or forests in the middle, and the lake at the bottom, respectively. Usually, images of closely viewed trees or flowers are highly textured; they exhibit no bias to a specific orientation. Images of city streets acquired in middle distances tend to contain highly textured areas (corresponding to trees, buildings, etc.) and smooth areas at the top (corresponding to the sky) and the bottom (corresponding to the ground). It has been shown that different categories of scenes (such as forest, mountain, city, beach, etc) tend to have different statistical properties in their spatial structures [10]. In addition, without detailed local information of objects, human subjects are able to identify scenes as long as the global relations between large-scale structures in the scenes are preserved [8]. These evidences indicate that spatial structures of scenes are probably used by humans to identify scene categories.

We propose an approach to categorize images into multiple classes by capturing their spatial structure characteristics. Instead of representing each image by a combination of widely used features such as color histogram, moments, edge direction, MSAR texture, etc., we derive the 2D Spatial Frequency Map (SFM), which characterizes the spatial structure of each image. To facilitate clustering, the dimensionality of the 2D SFM is reduced by orientation projection and principal component analysis. We use two spectral clustering methods and several similarity scores to study the effect of the SFMs.

The paper is organized as follows. The 2D spatial frequency map is introduced in §2. Low dimensional representations of SFM are discussed in §3. Clustering schemes are addressed in §4. Experimental results are discussed in §5.

2. SPATIAL FREQUENCY MAP

A scene image usually consists of multiple surfaces of different depths. The specific structure of the scene is reflected by the depth of these surfaces and their relative spatial relationships. Since spatial frequency analysis of the image provides depth information of the scene [4], we propose using the spatial distributions of local spatial frequency to characterize the scene structure that the image underlies.

We first divide each image ($M\times N$ pixels) into a set of square patches. Each patch has $N_P\times N_P$ pixels (in our setting, $0.05\times\min(M,N)<N_P<0.5\times\min(M,N)$). The distances between the centers of neighboring patches along horizontal and vertical directions are all N_R pixels. The overlapping of patches can be controlled by the parameters N_P and N_R. We then compute the 2D Fourier power spectrum of each patch. Boundary effects are reduced by applying a Kaiser-Bessel window [3], which has the maximum value at the center and gradually decreases to zero towards the boundary. Since it is isotropic, no orientation bias is introduced. The Fourier transform of each patch, $F_P(u,v)$, is given by:

$$F_P(u,v)=\sum_{(x,y)} I_{mn}(x,y)w(x,y)\exp[\frac{1}{N_P}2\pi i(ux+vy)] \tag{1}$$

where x and y are the horizontal and vertical coordinates of the image pixels, respectively; u and v are corresponding spatial frequency coordinates; $I_{mn}(x,y)$ is the image patch centered on (m,n); $w(x,y)$ is the Kaiser-Bessel window function.

For each image patch, the power spectrum $S_P(u,v)$ is given by,

$$S_P(u,v)=\frac{1}{N_P^2}\left|F_P(u,v)\right|^2 \tag{2}$$

The spatial frequency coordinates (u,v) can be represented by polar coordinates (f, φ), $u = f\cos\varphi$, $v = f\sin\varphi$, where f is the spatial frequency and φ the orientation. For a specific spatial frequency, by averaging the 2D power spectrum $S(f,\varphi)$ across all orientations, we obtain a 1D power spectrum $Q(f)$ as a function of spatial frequency f, i.e.,

$$Q(f) = \frac{1}{N_\varphi} \sum_\varphi S_P(f, \varphi) \tag{3}$$

where N_φ is the number of orientations.

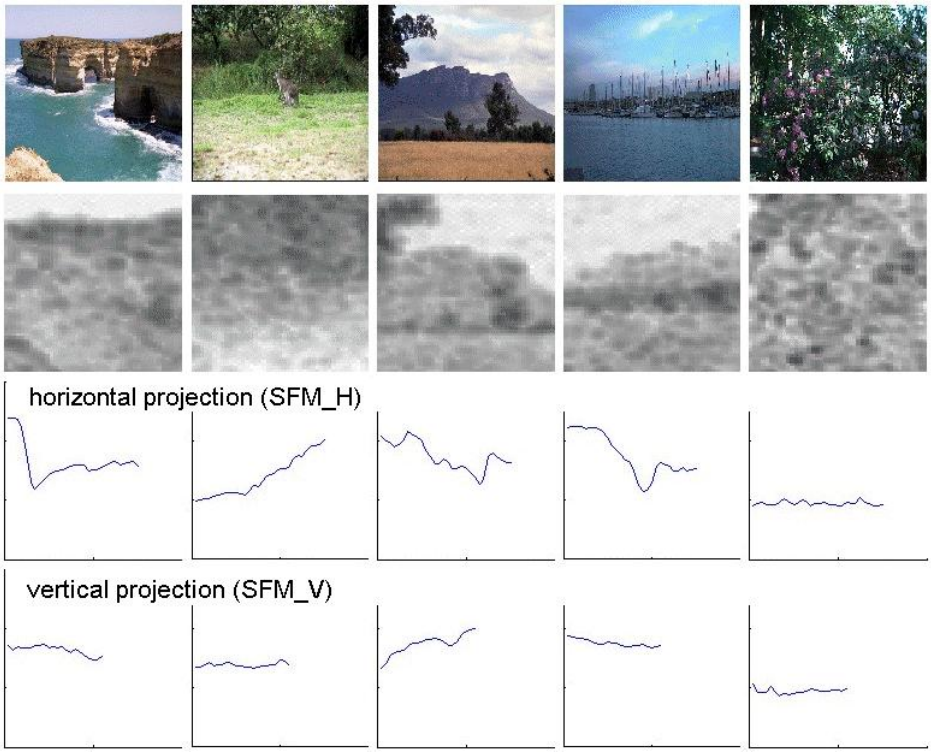


Figure 1. Spatial frequency maps (SFMs) and their projections. Upper row: original images. Each image is of 768×512 pixels. Second row: the corresponding SFMs. Each pixel in the SFM corresponds to a 21×21 patch in the original image. The distances between the centers of neighboring patches along horizontal and vertical directions are 3 pixels. The α values (see text) in the red, green, and blue channels are combined to a single α value, indicated by the intensity of the pixels in the second row. Third row: SFM_H (i.e. projecting SFMs along the horizontal direction). Bottom row: SFM_V (i.e. projecting SFMs along the vertical direction).

According to the power law, Fourier power decreases with the increase of the spatial frequency [12], i.e., $Q(f) \sim Af^{-\alpha}$ (where A is a constant determining the overall image contrast). Usually the function $Q(f)$ becomes a decreasing line (slope is α) when plotted in log-log scale. The value of α reflects how the power is distributed across the range of the spatial frequency. When α is small, the power decreases slowly as spatial frequency increases, thus an image patch contains a wide range of frequency components; when α is large, the power decreases quickly as spatial frequency increases, thus an image patch mainly contains low frequency information. Therefore,

we can use the parameter α to characterize the spatial frequency properties of each image patch. We estimate α via linear regression of $Q(f)$ plotted in the log-log scale.

Suppose an image is divided into $R \times T$ patches, we obtain an $R \times T$ matrix. Each element of the matrix is the α value of the corresponding patch. We call this matrix the 2D Spatial Frequency Map (SFM) of this image. SFM describes how the power is distributed across different spatial frequencies in local areas of the image. Some SFM examples are given in Figure 1.

3. LOW DIMENSIONAL REPRESENTATION OF THE SPATIAL FREQUENCY MAP

To classify each image into different categories, its corresponding SFM is regarded as a high-dimensional data point. The dimensionality equals the number of pixels in the SFM. For example, for a 150×200 image, if we divide it into square patches of 21×21 pixels and let the inter-patch distances be 3 pixels, then the dimension of SFM would be 2580 (i.e. 43×60). Since it is often difficult to cluster data points distributed sparsely in the high-dimensional space, we use two approaches, i.e. projection vector representation and principal component decomposition, to reduce the dimensionality of the SFM.

3.1 Projection Vector Representations

A simple way to reduce the dimensionality of SFM is to project the SFM along horizontal and vertical directions. Denote the element in the ith row and jth column of the SFM (with $R \times T$ elements) as α_{ij}. By projecting the SFM along the horizontal direction, we obtain a vector $\boldsymbol{z}_h = [b_1,\ldots,b_i,\ldots,b_R]'$, where $b_i = \Sigma_j \alpha_{ij}/T$. Similarly, by projecting the SFM along the vertical direction, we obtain a vector $\boldsymbol{z}_v = [c_1,\ldots,c_j,\ldots,c_T]'$, where $c_j = \Sigma_i \alpha_{ij}/R$. The horizontally and vertically projected vector can be concatenated into a long vector $\boldsymbol{z} = [\boldsymbol{z}_h ; \boldsymbol{z}_v]$. We denote the projection vectors of SFM along the horizontal and vertical directions as SFM_H and SFM_V, respectively. The concatenated vector is denoted as SFM_HV. The two lower rows of Figure 1 show examples of the SFM_H and SFM_V.

3.2 Principle Component Representations

The second approach we use to reduce the dimensionality of SFM is the principal component analysis (PCA). The main idea is to represent each SFM as the linear combination of a small number of principal components.

Suppose we have L images, and the SFM of each image contains G pixels (where $G = R \times T$). We combine all the SFMs and obtained an $L \times G$ matrix $\boldsymbol{D}$, where each row corresponds to the SFM of an image. To compute the principal components of the SFMs, we first calculate the covariance matrix of the $\boldsymbol{D}$, i.e. $\boldsymbol{C} = \text{cov}(\boldsymbol{D})$, which is a $G \times G$ symmetric real matrix. It can be proved that $\boldsymbol{C}$ has G real eigenvalue/eigenvector pairs: $\boldsymbol{C} \times \boldsymbol{V} = \boldsymbol{V} \times \boldsymbol{\lambda}$, where each column of $\boldsymbol{V}$ is an eigenvector, $\boldsymbol{\lambda} = \text{diag}(\lambda_1,\ldots,\lambda_i,\ldots,\lambda_G)$ and λ_i is the ith eigenvalue. Then, we select the K eigenvectors (principal components) corresponding to the K-largest eigenvalues. These K principal components of the SFMs constitute a $G \times K$ matrix, denoted as $\boldsymbol{E}_K$.

Next, each G-dimensional SFM point is projected onto the K-dimensional space of the K principal components. This is achieved by multiplying $\boldsymbol{D}$ and $\boldsymbol{E}_K$, i.e., $\boldsymbol{D}_K = \boldsymbol{D} \times \boldsymbol{E}_K$. The resultant $L \times K$ matrix $\boldsymbol{D}_K$ represents L low-dimensional SFMs. Because K is usually much smaller than G, $\boldsymbol{D}_K$

is a much more compact representation of the original data. It is likely that the clusters in this low dimensional space become more distinguishable than in the original high dimensional space.

Because the variances of various dimensions can differ significantly, different dimensions will play significantly different roles in SFM-comparison if their effect is not normalized. For example, if the first eigenvalue λ_1 is 10 times of the second eigenvalue λ_2, the difference along the first eigenvector direction is about 10 times of that along the second eigenvector direction. This makes some common similarity measures (e.g. L_1 or L_2 similarity, or the correlation coefficient) inapplicable to clustering the data $\boldsymbol{D}_K$. We overcome this problem via whitening the data, so that for each eigenvector-direction, the mean value is 0, and the variance is 1. Without confusion, we also call the whitened data $\boldsymbol{D}_K$.

4. CLUSTERING METHODS

In terms of the SFM representations (SFM_H, SFM_V, SFM_HV, or principal component representation), each image is represented as a data point in the respective space. In general, we denote the dimensionality of the space as K. We cluster these data points using the following methods.

4.1 Similarity Scores

We use three similarity scores of data points. First, we consider the L_1 similarity, or S_{L1}, which is defined as the exponent of the negative L_1 distance (thus the range is [0, 1]).

Second, we consider the correlation coefficient similarity, S_{CC}, of the L data points. The correlation coefficients are linearly normalized to the range [0, 1]. Different from S_{L1}, S_{CC} measures the "similarity" of the "variations" of data samples.

Third, we combine both the L_1 similarity and the correlation coefficient similarity, i.e. $S_{L1+CC} = 0.5\times(S_{L1} + S_{CC})$.

4.2 Spectral Clustering Method

Recently the spectral clustering approach becomes popular [7][9][2]. This approach uses the top eigenvectors of the similarity matrix as the indicators of different clusters. Compared to other clustering approaches, it can generate globally more coherent clusters. In this paper, we use the MinMaxCut spectral clustering method [2]. MinMaxCut uses the second smallest eigenvector of the similarity matrix, called Fiedler vector, as the indicator of data bipartition. Multiple clusters are obtained by iteratively using the MinMaxCut. It has been proved that MinMaxCut favors balanced cut, so that the resultant clusters are likely to have comparable sizes.

4.3 MST-Spectral Clustering Method

One potential problem of spectral clustering is that since low similarity values tend to be noisy, a noise-removal method is usually needed to obtain reliable clusters. As an effort to improve the clustering method, we introduce the Maximal Spanning Tree (MST) into spectral clustering [6]. More specifically, instead of using spectral clustering directly on the similarity matrix, we first construct the MST from the similarity matrix and then use spectral clustering. Because MST typically preserves the cluster information of data (e.g. [14]), we expect this MST-spectral clustering approach would generate good clusters, given that the clusters do exist in the data. Since MST can be described as an extremely sparse matrix, an additional advantage of MST-spectral

clustering is that very large number of data points (and accordingly very large similarity matrix) can be handled more easily.

To construct the MST from the similarity matrix, we treat each data point as a graph node. The similarity value of each pair of nodes is set as the weight of the respective edge between these nodes. We use the Prim's algorithm [1] to construct MST from the fully connected graph. Since the number of undirected edges is $(L^2-L)/2$, the MST construction has the complexity $O((L^2-L)/2|\lg L)$ when the ordinary binary heap is used and can be reduced to $O((L^2-L)/2+L\lg L)$ when the Fibonacci heap [1] is used. The constructed MST is an undirected graph; the respective adjacency matrix, denoted as $\boldsymbol{T}$, is symmetric.

We add an $L\times L$ identity matrix $\boldsymbol{I}_L$ to $\boldsymbol{T}$ to generate a symmetric matrix $\boldsymbol{S}$, i.e. $\boldsymbol{S} = \boldsymbol{T} + \boldsymbol{I}_L$, hence the self-similarity of every data point is the strongest. Because $\boldsymbol{S}$ is a symmetric real matrix, it has L real eigenvalue-eigenvector pairs. This indicates spectral clustering can be applied to this MST-based similarity matrix.

We note that within the framework of MST-spectral clustering, other spectral clustering methods (e.g. normalized cut [9], etc) can be used as well.

4.4 Clustering Results Evaluation Method

The clustering results are evaluated using a score similar to the F-measure used in information retrieval [13]. Denote the image set of the ith cluster as C_i. Denote the ith ground truth image-cluster as $\hat{C}_i$ (obtained by asking human subjects to categorize images into different categories). We calculate the proportion of images in the clustering results that are correctly put into the jth ground truth cluster, i.e., $\eta_{ij} = |C_i \cap \hat{C}_j|/|C_i|$. We also compute the proportion of mis-clustered images that should belong to the jth ground truth cluster $\xi_{ij} = |C_i \cap \hat{C}_j|/|\hat{C}_j|$. The score to measure clustering quality of the ith category, still called F, is then defined as:

$$F = \left\langle \max_j \left(\frac{\eta_{ij}\xi_{ij}}{\eta_{ij}+\xi_{ij}} \right) \right\rangle_i \tag{4}$$

5. EXPERIMENTS

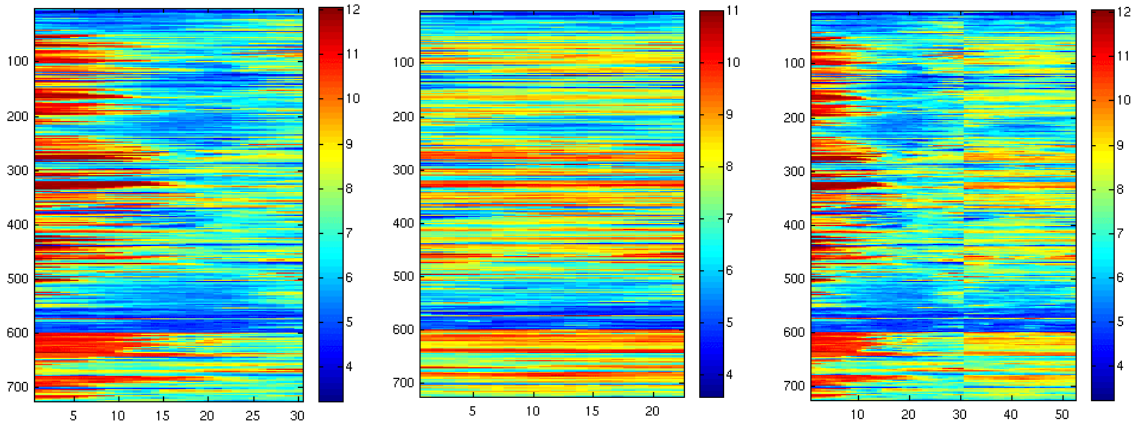
As a preliminary study, we investigated our approach using a database of 724 images of natural scenes from University of Washington[*]. It includes (1) images of trees, flowers, and bushes in short distance, (2) city and campus sceneries in middle distance, and (3) lake, sea, grassland, and mountains in long distance. The task is to automatically categorize these images based on spatial frequency maps.

For all images, we first generated the SFMs (T=30, R=22. See some examples in the second row of Figure 1). Then, both the project vector representations (Figure 2) and principal component representations (Figure 3) were produced.

In Figure 2, the horizontal, vertical, and the concatenated projection vectors are shown. It is obvious that the horizontal projection vectors of different images (i.e. the different rows in (a)) have rich differences in both mean values and variances, indicating both the L_1 and correlation coefficient similarities should have discriminative strength in differentiating these images. In contrast, for the vertical projection vectors in (b), the mean values tend to be more distinctive than

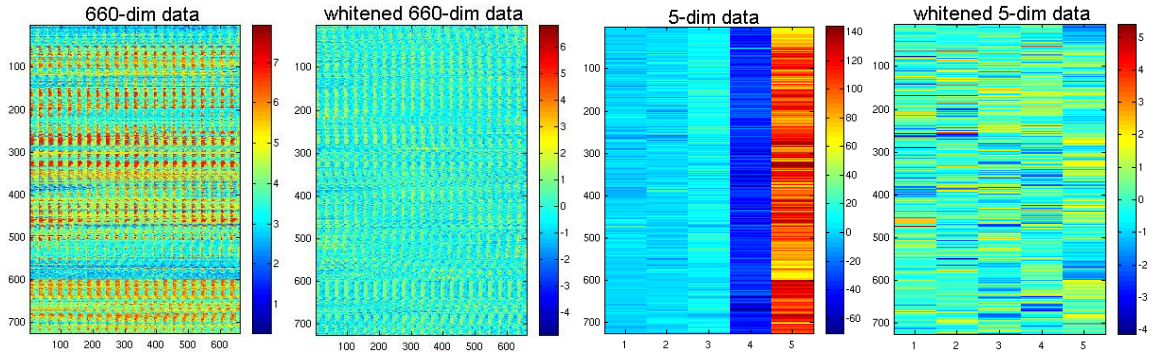
[*] http://www.cs.washington.edu/research/imagedatabase/groundtruth

the variances, suggesting that the correlation coefficient similarity would not have the same discriminative strength as the L_1 similarity. The concatenated projection vectors in (c) combine both SFM_H and SFM_V. Visually, the S_{L1} and S_{CC} similarity scores should have close discriminative strength. (These observations will be verified in later part of this section).



(a) *SFM_H* (b) *SFM_V* (c) *SFM_HV*

Figure 2. Projection vector representations of 724 natural scene images. For each plot, the rows correspond to the projection vectors of the images, and the columns correspond to the bins of projection.



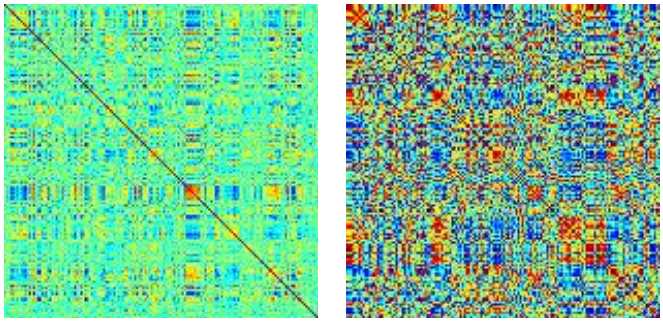
(a) *Original data* (b) *Dimension-reduced data*

Figure 3. Principal component representation of SFMs. The original data $\boldsymbol{D}$ *(left-most) is represented as a 724×660 array; lower-dimensional data* $\boldsymbol{D}_K$ *are generated via principal component decomposition (an example of K=5 is shown). For the low-dimensional data, from-right-to-left the columns correspond to decreasing eigenvalues (e.g. the right most columns correspond to the largest eigenvalue).*

In Figure 3, we show an example of the principal component representations of SFMs. The dimensionality of input SFM data is 660 (=30×22), as shown in (a). It is apparent different images have different scales, indicating that for this original SFM data, L_1 similarity would have more discriminative power than the correlation coefficient similarity. Because this is little visual variation among columns of (a), we can use the principal components to reduce the dimensionality. In (b), we show the decomposition using the five top (largest) eigenvectors. The new data $\boldsymbol{D}_K$ (K=5) have different variances along different eigenvector directions. After data whitening, the variances are comparable, and presumably the data $\boldsymbol{D}_K$ can be more effectively clustered.

For simplicity, in the following we use S_{660} to denote the similarity matrixes using the original 660-dimensional SFM data, and S_5 for the 5-dimensional principal component representations. They were calculated using three similarity scores, S_{L1}, S_{CC}, or S_{L1+CC}. In Figure 4, we show the similarity matrix using S_{CC}. It can be seen that after dimension-reduction, the contrast of similarity /dissimilarity of images are clearer.

The image clustering can be understood as reordering the rows/columns in the similarity matrix so that similar images are grouped together and dissimilar images are separated. Accordingly, if a cluster of images is coherent, the pair-wise similarity value will be large; if two clusters are distinctive, their mutual similarity values will be low. In terms of the similarity matrix, good clustering results mean that the respective diagonal sub-matrix of a coherent cluster has large similarity values (most red-colored), and the off-diagonal sub-matrix will have small similarity values (most blue-colored).



(a) S_{660} (b) S_5

Figure 4. The correlation coefficient similarity matrixes of the SFMs and the dimension-reduced data. Red color means higher similarity, and blue color means lower similarity.

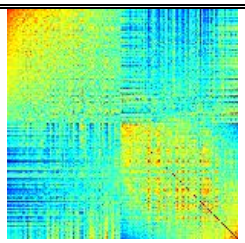
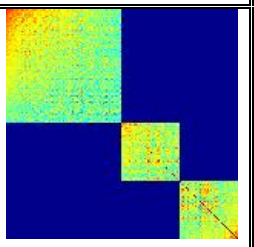
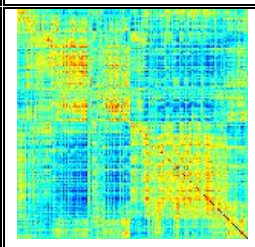
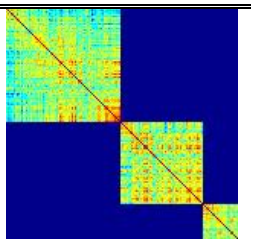
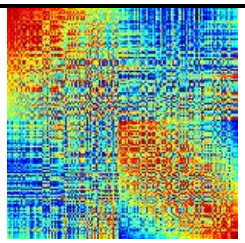
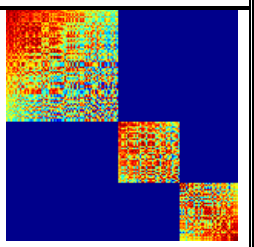
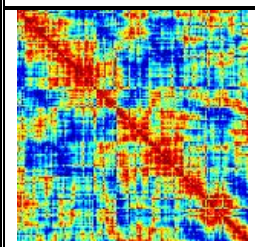
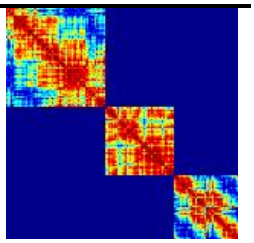
Input S	Spectral clustered		MST-spectral clustered	
	Reordered S	In-cluster portion of the reordered S	Reordered S	In-cluster portion of the reordered S
S_{660}				
S_5				

Figure 5. Illustration of clustering schemes for similarity matrixes in Figure 5. The 3 major clusters were obtained using the MinMaxCut spectral clustering. Compared to the spectral-clustering, MST-spectral clustering generates more coherent clusters, so that the diagonal sub-matrixes are more densely red-colored, and the off-diagonal sub-matrixes are more densely blue-colored.

(a) *Examples of the first image-cluster, corresponding to short-distance scenes*

(b) *Examples of the second image-cluster, corresponding to middle-distance scenes*

(c) Examples of the third image-cluster, corresponding to long-distance scenes

Figure 6. An example of clustering results. 5-dimensional PCA representations of SFM and the MST-spectral clustering are used. The target number of clusters is set to be 3.

Figure 5 shows a clustering example using the S_{CC} similarity matrix. In the reordered similarity matrixes and the respective in-cluster portions, the clustering results of S_5 are better than those of S_{660}. This suggests that dimension-reduction plays a very useful role. Figure 5 also indicates that the clusters produced by MST-spectral clustering method (for S_{CC} of S_5) are coherent.

In Figure 6, we show example images of the detected clusters. The results show that in the first category (see (a)), most images are highly textured; they belong to close-viewed scenes. In the second category (see (b)), most images contain both objects in short distance (such as trees and buildings close to the camera) and long distance (such as the sky). Thus, some areas (usually in the upper part of the image) in the images are quite smooth, whereas others (usually in the middle or

lower part of the image) are highly textured. In the third category (see (c)), most images contain horizontally distributed smooth areas and are images of long distance scenes. This example indicates that SFM does capture the distribution of the spatial frequency of different areas in an image, thus reflecting different spatial layout and depth of the scene.

Table 1. F-measures for clustering projection vector representations.

Similarity	SFM_H		SFM_V		SFM_HV	
	Spectral	MST-spectral	Spectral	MST-spectral	Spectral	MST-spectral
S_{L1}	0.5550	0.5368	0.5730	0.5620	0.5548	0.5584
S_{CC}	0.5234	0.4095	0.4436	0.4307	0.5235	0.5104
S_{L1+CC}	0.5454	0.5365	0.5015	0.4874	0.5398	0.5411

Table 2. F-measures for clustering PCA representations.

Similarity	S_{660}		S_5	
	Spectral	*MST-spectral*	*Spectral*	*MST-spectral*
S_{L1}	0.5591	0.5486	0.5406	0.5423
S_{CC}	0.4823	0.4747	0.5062	0.5491
S_{L1+CC}	0.5412	0.5499	0.5392	0.5404

We then systematically compared different feature representations, similarity scores, and clustering schemes using the ground truth clusters determined by human-subjects. Table 1 shows the results of clustering projection vector representations. The *F*-measure was used: the larger the score, the better the consistency between the clustering results and the ground truth clusters. We have the following observations. (1) For horizontal projection vectors (SFM_H), S_{L1}, S_{CC}, and S_{L1+CC} have comparable discriminative strength. (2) For the vertical projection vectors (SFM_V), S_{L1} is much more discriminative than S_{CC}. (3) It is a safe choice to combine both horizontal and vertical projections as SFM_HV. The respective *F*-measures are good for all the three similarity scores. (4) MST-spectral clustering and the spectral-clustering have comparable performances for the projection vector representations. We note that the results in Table 1 are consistent with our observations of Figure 2.

Table 2 compares the clustering results of principal component representations. (1) For both the original SFM data S_{660} and the 5-dimensional data S_5, the L_1 similarity is much more discriminative than the correlation coefficient similarity (e.g. the *F*-measure is 0.55 versus 0.48 for S_{660}). This is consistent with our observation on Figure 3. (2) MST-spectral clustering is (slightly) better than the spectral clustering for S_5. (3) The combination of L_1 and correlation coefficient similarities is a safe choice for good clustering results.

The overall results of Tables 1 and 2 suggest that for various feature representations and clustering schemes, the L_1 similarity of SFMs is good in discriminating different images. The *F*-measure is always larger than 0.54.

To investigate how good these clustering results are, we generated 3 random clusters (but with approximately the same numbers of images in each cluster). The average *F*-measure of 10 trials is 0.38. Because 0.54 is significantly larger than 0.38, we have confidence that the SFM does play a positive role in discriminating the scene images in our experiments.

REFERENCES

[1] Corman, T.H., Leiserson, C.E., and Rivest, R.L., *Introduction to Algorithms*, MIT, 2001.

[2] Ding, C., He, X., Zha, H., Gu, M., and Simon, H., "A min-max cut algorithm for graph partitioning and data clustering," *Proc. of 1st IEEE Int'l Conf. Data Mining*, San Jose, CA, pp.107-114, 2001

[3] Harris, F.J. "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proceedings of the IEEE*, 66, pp.51-83, 1978.

[4] Kakai, K., and Finkel, L.H., "Spatial-frequency analysis in the perception of perspective depth," *Network: Comput. Neural Syst.*, 8, 335-352, 1997.

[5] Luo, J. and Savakis, A., "Indoor vs outdoor classification of consumer photographs using low-level and semantic features," *2001 Int Conf on Image Processing*, pp.745-748, 2001.

[6] Peng, H.C., He, X., and Long, F., "Automatic content extraction of filled form images based on clustering component block projection vectors," *IS&T/SPIE 16th Annual Symp. on Electronic Imaging, Conf. on Document Recognition and Retrieval XI*, San Jose, CA, USA, 2004.

[7] Pothen, A., Simon, H.D., and Liou, K.P., "Partitioning sparse matrices with eigenvectors of graph," *SIAM Journal of Matrix Anal. Appl.*, 11, pp.430-452, 1990.

[8] Schyns, P.G., and Oliva, A., "From blobs to boundary edges: evidence for time and spatial scale dependent scene recognition," *Psychol. Sci.*, 5, pp.195-200, 1994.

[9] Shi, J., and Malik, J., "Normalized cuts and image segmentation," *IEEE Trans. PAMI*, 2000.

[10] Torralba, A., and Oliva, A, "Statistics of natural image categories," *Network: Comput. Neural Syst.*, 14, pp.391-412, 2003.

[11] Vailaya, A., Figueiredo, M.A.F., Jain, A.K., and Zhang, H.J., "Image classification for content-based indexing," *IEEE Trans. on Image Processing*, 10(1), pp.117-130, 2001.

[12] van der Schaaf, A., and van Hateren, J.H., "Modelling the power spectra of natural images: statistics and information," *Vision Research*, 36(17), pp. 2759-2770, 1996.

[13] van Rijsbergen, K. *Information Retrieval* (2nd Ed.), (www.dcs.gla.ac.uk/Keith/Preface.html) Butterworths, London, 1979.

[14] Xu, Y., Olman, V., and Xu, D., "Clustering gene expression data using a graph-theoretic approach: an application of minimum spanning tree," *Bioinformatics*, 18(4), pp.526-535, 2002.

[15] Zhang, H.J. and Zhong, D., "A scheme for visual feature based image indexing," *Proc. SPIE Conf. on Storage and Retrieval for Image and Video Databases*, pp. 36-46, San Jose, CA, 1995.